Feature Embedding Based Text Instance Grouping for Largely Spaced and Occluded Text Detection

Pan Gao, Qi Wan, RenWu Gao and LinLin Shen*
Computer Vision Institute
School of Computer Science and Software Engineering
Shenzhen University, Shenzhen, China
Email: {gaopan2017, wanqi2019}@email.szu.edu.cn, {re.gao, llshen}@szu.edu.cn

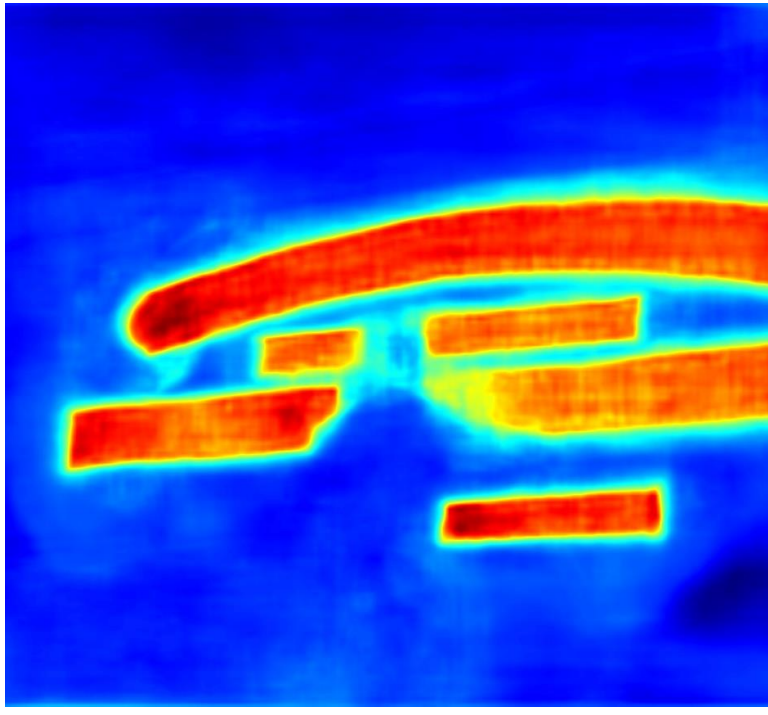
The “Over Segmentation” Problem

Text with large space between texts/characters

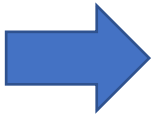
Text with partial occlusion


ΦΡΟΥΤΕΜΠΟΡΙΚΗ
ΣΑΝΤΟΡΙΝΗΣ
ΤΗΛ. 22884

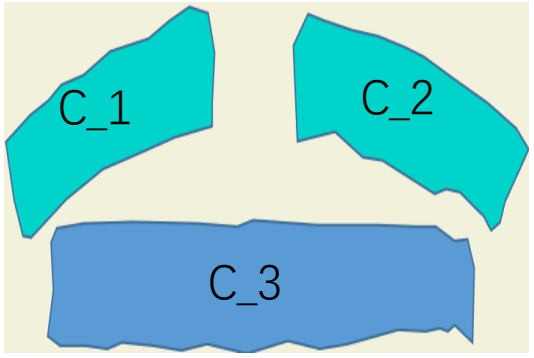

ΤΗΛ. 22884

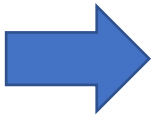


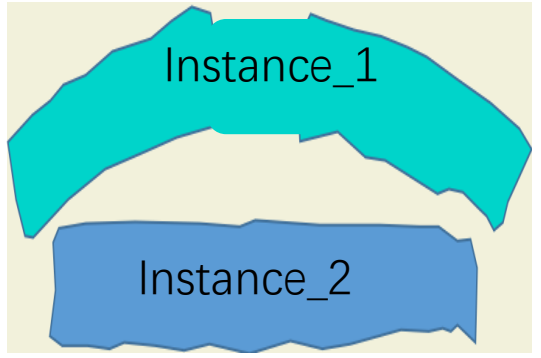

HIGH POINT
FURNITURE



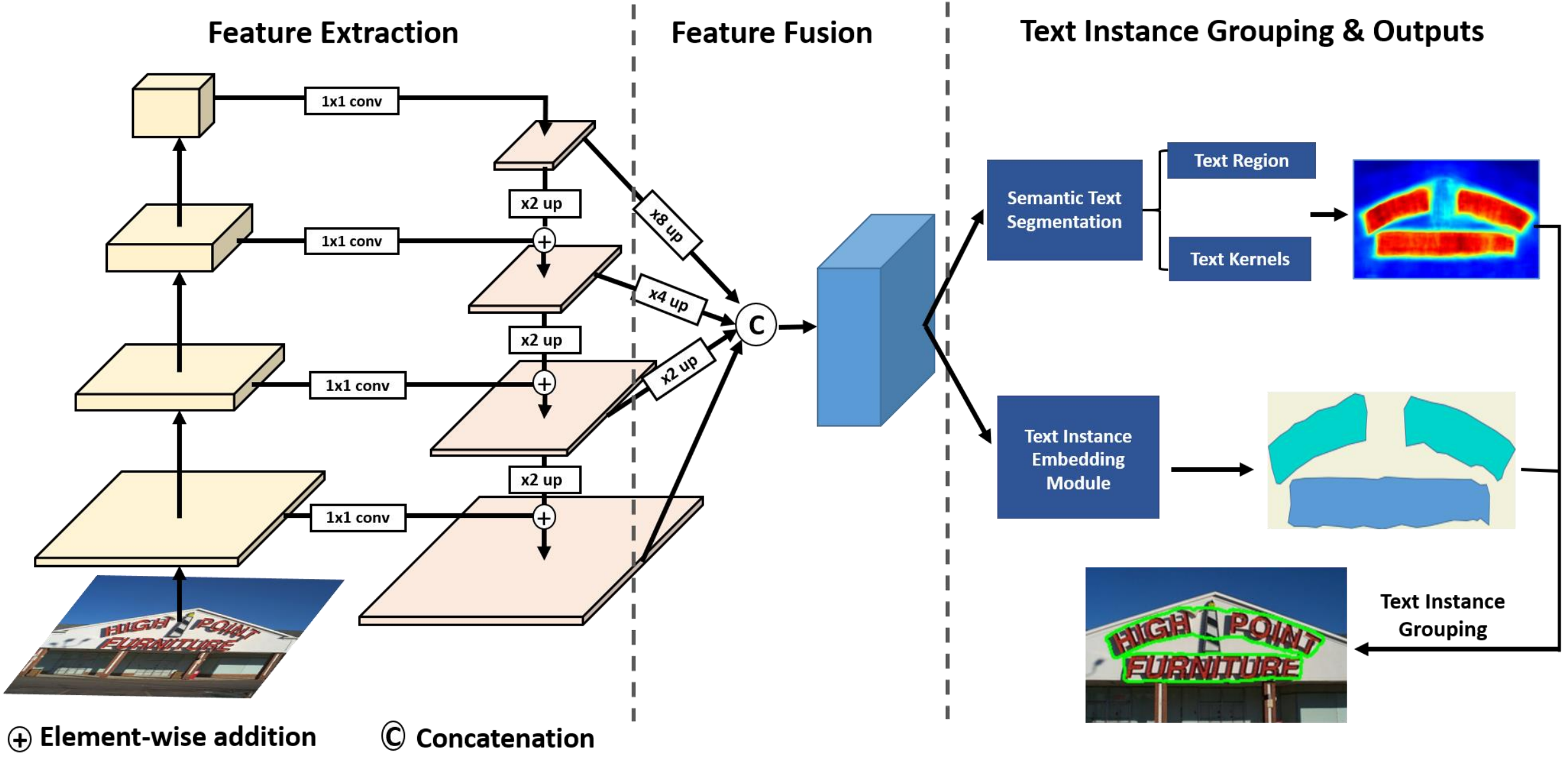

HIGH POINT
FURNITURE


C_1
C_2
C_3



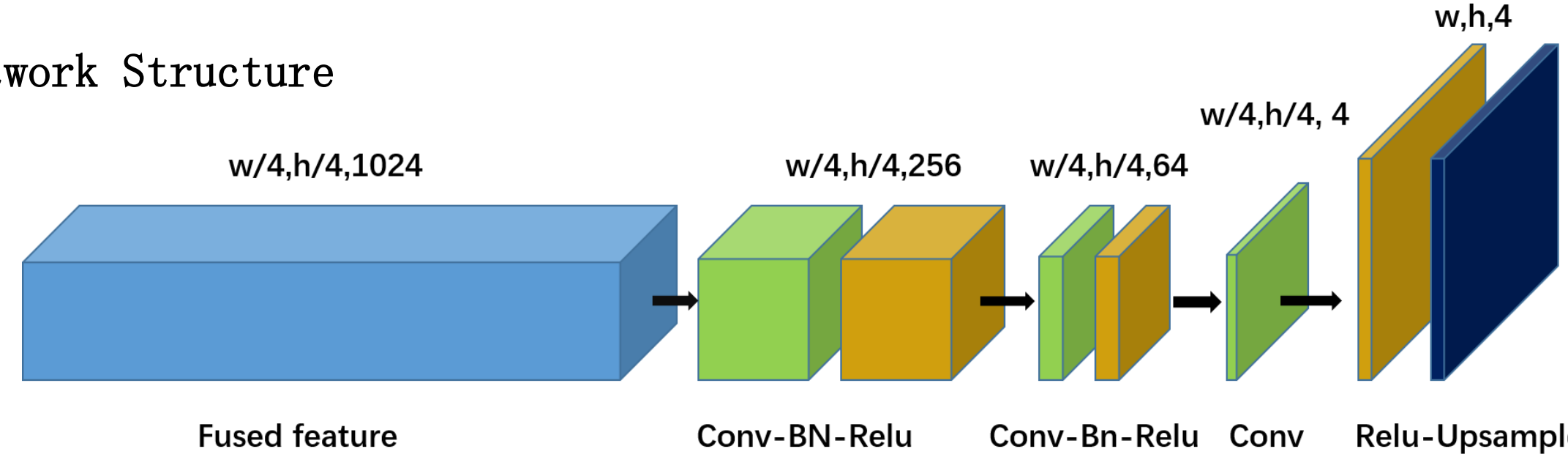

Instance_1
Instance_2

The pipeline of our framework



The Details Of Text Instance Embedding Module

Network Structure



Loss Function

a) *Reducing the intra-instance distance*

$$L_{intra} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|O(T_i)|}\sum_{p\in O(T_i)}\ln(Dis(p,T_i)+1)$$

$$Dis(p,T_i) = max(\|F_p - F_{T_i}\| - \theta_{intra}, 0)^2$$

b) *Increasing the inter-instance distance*

$$L_{inter} = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1 j\neq i}^{N}\ln(Dis(K_i,K_j)+1)$$

$$Dis(K_i,K_j) = max(\theta_{inter} - \|F_{K_i} - F_{K_j}\|, 0)^2$$

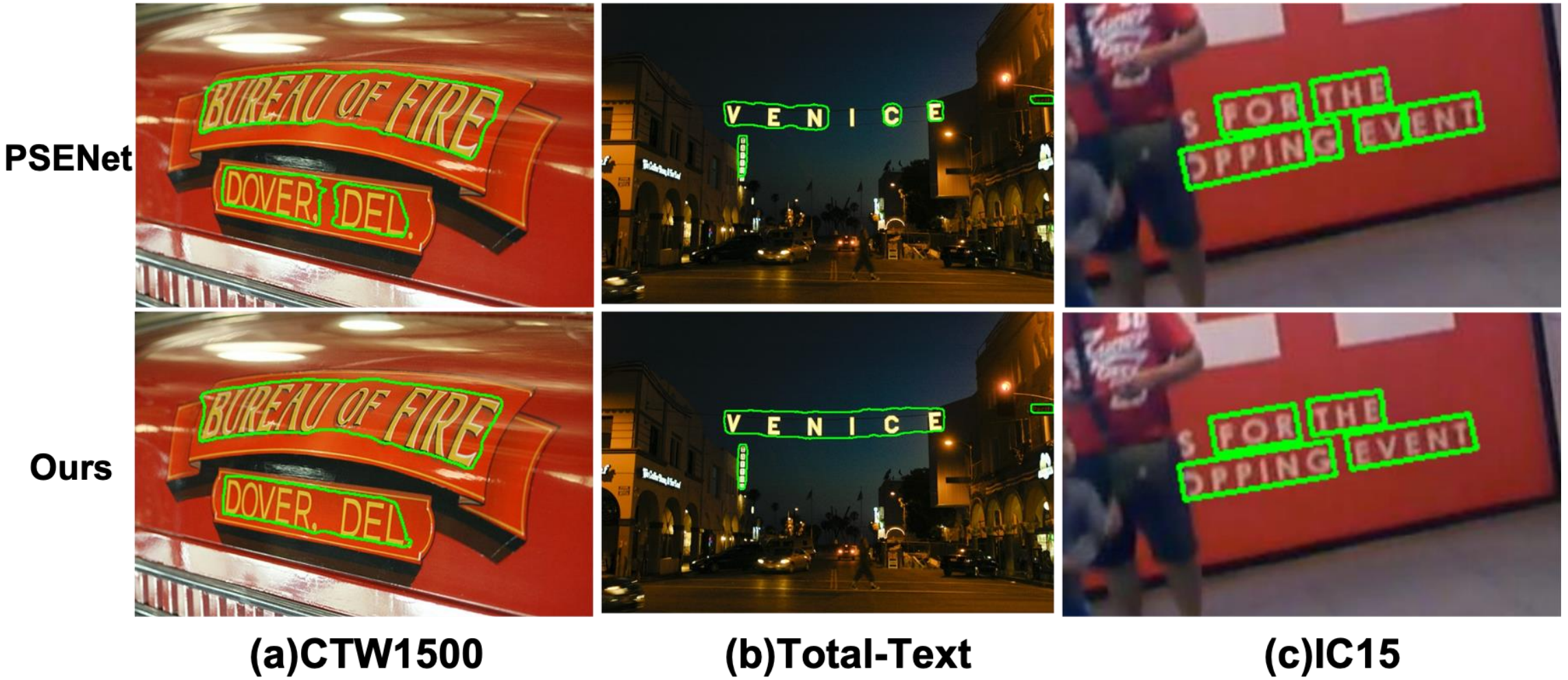
Ablation Study

(a)Before grouping (b)After grouping

Table.1 *Ablation Study* on CTW-1500.

Method	CTW-1500		
	P	R	F
baseline(reimplement PSENet)	80.6	75.6	78.1
baseline + TIEM	83.6	77.6	80.5
baseline + TIEM + instance grouping	**85.1**	**77.9**	**81.3**

Experiment Result



Example results of our method on CTW-1500(a), Total-Text(b) and IC15(c).

Table.2 Single-scale Result On CTW-1500. “P”, “R” And “F” Represent The Precision, Recall And F-measure Respectively.

Method		CTW-1500		
		P	R	F
Training from scratch	CTPN*[23]	60.4*	53.8*	56.9*
	Seglink* [24]	42.3*	40.0*	40.8*
	EAST*[7]	78.7*	49.1*	60.4*
	CTD+TLOC [19]	77.4	69.8	73.4
	PSENet-1s [15]	80.6	75.6	78.0
	PAN-640 [1]	84.6	77.7	81.0
	Ours	**85.1**	**77.9**	**81.3**
Pre-trained on external dataset	TextSnake[25]	67.9	**85.3**	75.6
	CSE[26]	78.7	76.1	77.4
	LOMO[27]	**89.2**	69.6	78.4
	SAE[2]	82.7	77.8	80.1
	TextField[13]	83.0	79.8	81.4
	MSR[14]	84.1	79.0	81.5
	PSENet-1s[15]	84.8	79.7	82.2
	DB[28]	86.9	80.2	83.4
	CRATF[29]	86.0	81.1	83.5
	PAN-640[1]	86.4	81.2	**83.7**
	Ours	87.9	79.9	**83.7**

Table.3 Single-scale Result On Total-Text. “P”, “R” And “F” Represent The Precision, Recall And F-measure Respectively.

Method		Total-Text		
		P	R	F
Training from scratch	Seglink*[24]	30.3*	23.8*	26.7*
	EAST*[7]	50.0*	36.2*	42.0*
	DeconvNet [20]	33.0	40.0	36.0
	PSENet-1s [15]	81.8	75.1	78.3
	PAN-640 [1]	**88.0**	79.4	**83.5**
	Ours	82.8	**79.6**	81.2
Pre-trained on external dataset	TextSnake[25]	82.7	74.5	78.4
	ATTR [30]	80.0	76.2	78.5
	MSR[14]	85.2	73.0	78.6
	CSE[26]	81.4	79.7	80.2
	TextField[13]	81.2	79.9	80.6
	PSENet-1s[15]	84.0	77.9	80.9
	LOMO[27]	88.6	75.7	81.6
	CRAFT[29]	87.6	79.9	83.6
	DB[28]	87.1	**82.5**	84.7
	PAN-640[1]	**89.3**	81.0	**85.0**
	Ours	87.0	79.3	83.0

Table.4 Single-scale Result On IC15. “P”, “R” And “F” Represent The Precision, Recall And F-measure Respectively.

Method		IC15		
		P	R	F
Training from scratch	RRPN[6]	82.0	73.0	77.0
	EAST[7]	**83.6**	73.47	78.2
	DeepReg[31]	82.0	80.0	81.0
	CTPN[23]	82.0	73.0	77.0
	PAN-736[1]	82.9	77.8	80.3
	PSENet-1s[15]	81.5	79.1	80.9
	PixelLink[11]	82.9	**81.7**	**82.3**
	Ours	83.4	80.7	82.1
Pre-trained on external dataset	SSTD[32]	82.2	73.9	76.9
	SegLink[24]	73.1	76.8	75.0
	WordSup[33]	79.3	77.0	78.2
	Lyu et al.[12]	**94.1**	70.7	80.7
	RRD [5]	85.6	79.0	82.2
	TextField[13]	83.3	80.5	82.4
	TextSnake[25]	84.9	80.4	82.6
	PAN-736[1]	84.0	81.9	82.9
	PSENet-1s[15]	86.9	84.5	85.7
	SAE[2]	88.3	**85.3**	86.6
	CRATF[29]	89.8	84.3	86.9
	LOMO[27]	91.3	83.5	87.2
	DB-1152[28]	91.8	83.2	**87.3**
	Ours	87.1	80.0	83.4

THANKS

Email: gaopan2017@email.szu.edu.cn